

## **Chapter 5: CONCLUSIONS AND RECOMMENDATIONS**

### **Updated Conclusions**

In our June 30 report, we concluded that a great deal of progress had been made in defining and developing the HSEE and that results to date were quite positive. We also pointed out that much needed to be done before the HSEE could be administered operationally and that there was good reason to be concerned that many students are not well prepared to succeed on this examination. Nothing in our supplemental analyses leads us to change these conclusions.

As was the case for the multiple-choice questions, the quality of the essay questions was found to be high. The majority of the essay questions for reading passed the statistical screens, as did all of the writing scores. The essay questions appear to function in the same way for both males and females and for Hispanic and white students. Field test sample sizes were too small, however, to test for equivalent functioning in other groups.

The average scores for the essay questions were consistent across prompts, but generally low. Few students received maximum credit for their responses to these questions. The finding that students did not do well on these questions was no worse than the low levels of performance on the multiple-choice questions. We note, however, that while almost all students attempted to answer all of the multiple choice questions, a significant number of students (at least 10%) did not respond to both essay prompts. Some students may not have been highly motivated to do well on these questions. Scores may be higher when test results count, but it is not possible to estimate how much difference this would make. Even among students who did respond fully, fewer than half received 3 or 4 of the 4 possible points.

In Chapter 3, we presented passing rates and curriculum-alignment information for each of the proposed content standards. In mathematics, both passing rates and curriculum-alignment ratings were low for the Algebra 1 standards. This should not be surprising since students are not now required to take algebra in many districts and algebra is not mandated in state legislation. In English/language arts, passing rates and curriculum-alignment ratings were generally lower for standards involving higher levels of integration and analysis. There were few, if any, content standards for either test on which the majority of students performed well.

The results from the Teacher and Principal Surveys are largely the same as those presented in our June 30 report. There was considerable variation across the 40 to 50 participating schools in the nature and extent of current preparation for the HSEE and in expectations for its impact on the schools and their students.

In the next two sections of this chapter, we offer suggestions and recommendations concerning the large amount of work remaining to be done in developing the HSEE. We also offer suggestions for responding to concerns that students are currently not well prepared to pass the HSEE. These sections are followed by more specific recommendations on technical issues based on our review of the test developer's plans and proposals and then further discussion of the specific recommendations in our June 30 report.

### **Action Plans for a Tight Timeline**

The overarching recommendation in our June 30 report was that the Legislature, Governor, Board, and Department consider delaying full implementation of this new graduation requirement. Since the timeline for implementation has been written into the California Education Code, the primary audience for this recommendation was the Legislature. Superintendent Eastin and the Board have pointed out, quite correctly, that they have no choice but to follow the law in this matter. So far, there is no indication that the Legislature will consider changes prior to initial implementation in 2001.

We continue to be concerned about the timeline for implementing this graduation requirement for two reasons. The first concern is that, on the proposed schedule, the net outcome for students may well be negative. Based on our assessment of current student achievement relative to the standards likely to be adopted, it seems probable that a significant number of students in the first few high school classes subject to this requirement will not be adequately prepared. Without adequate preparation, the number of students who are denied diplomas will increase sharply as will the number of students who lose hope and drop out. The consequences for these students will necessarily be quite negative.

The Board, the HSEE Panel and others expressed concerns that, because of our initial recommendation, schools might relax efforts to implement programs to teach the state content standards. Schools please note: It is not likely that this program will be delayed and you have a lot to do. Please do not interpret either of our reports as a reason to back off from full and rapid implementation of the state content standards.

If the concern is with what schools should do, we suggest it would be more reasonable to begin by establishing a system of consequences for schools and give that system some time to work before enforcing consequences for students. The Academic Performance Index (API) is designed as part of just such a school-consequence system. Schools receive rewards or sanctions based on student performance as measured by the API. To date, however, the API only includes Stanford 9 scores. The API does not yet include measures of student achievement relative to the state content standards. It would be reasonable to move to include standards-based measures in the API as soon as possible and hold schools accountable for student performance on these measures before implementing important student consequences.

Our second concern is that, because timelines are so tight, implementation without a full dress rehearsal may well be flawed and the whole program then rejected before it can achieve its desired effects. If initial passing rates are quite low, there is likely to be a significant political backlash against the program. In particular, the experience in other states suggests a reasonable probability that the fate of the HSEE will be decided by the courts. Chapter 2 of our June 30 report identified several legal concerns, the most critical of which is the need to demonstrate that all students are provided adequate instruction in the material covered by the exam. (Note that “adequate” is likely to mean adequate quality as well as adequate content.) It seems risky, at best, to implement the HSEE requirement without first assembling evidence that schools are providing adequate instruction on the content of the test. Under California law, suits are likely to be directed at schools and districts as well as at the state. Individual

districts need to be given time and assistance to assemble evidence supporting the adequacy of their standards-based instruction.

Based on these concerns, our first recommendation continues to be:

*Recommendation 1. The Legislature and Governor should give serious consideration to postponing full implementation of the HSEE requirement by 1 or 2 years.*

We recognize that the primary task for the independent evaluation is to report on consequences of the HSEE, not necessarily to try to predict them in advance. Balancing risks associated with implementing a program too quickly with risks associated with implementing too slowly requires policy decisions. Such decisions are appropriately made by the Legislature, Governor, Board, and Superintendent who are charged with responsibility for setting such policy. The present implementation schedule for HSEE reflects the current judgment of these groups.

An important role for the independent evaluation that is clearly within the scope of our charge is to make recommendations for improving the development and implementation of the HSEE. In the remainder of this section, we assume that the current schedule will be followed and offer suggestions for maximizing the quality and minimizing the risks of the HSEE program under the constraints implied by this schedule. Most of these suggestions would also be appropriate should a delay be approved.

The co-chairs of the HSEE Panel suggested that it would be helpful to follow the discussion of issues experienced by other states (in Chapter 2 of our June 30 report) with a suggested agenda or plan of action for addressing these issues. In discussing our recommendation to postpone implementation, we did list many of the critical steps that need to be completed. We did not, however, suggest specific responsibilities and timelines for these steps and our list was more suggestive than comprehensive. As evaluators, there are limits to the extent to which we believe we should be responsible for creating, as well as evaluating, HSEE development plans. Nonetheless, we are in full agreement that a more detailed and public plan of action for addressing issues in implementing the HSEE is needed. We therefore offer the following recommendation:

*Recommendation 2. CDE should develop and seek comment on a more detailed timeline for HSEE implementation activities. This timeline should show responsibility for each required task and responsibility for oversight of the performance of each task. The plan should show key points at which decisions by the Board or others are required along with separate paths for alternative decisions that may be made at each of these points.*

Since the June 30 report, CDE has made considerable progress in planning the remaining steps that must be completed prior to implementation of the HSEE. They have received and reviewed AIR's timeline for the fall field test and the development of initial forms, set a date for convening a panel to recommend a minimum passing score, and worked with the HSEE Panel in their development of recommendations for accommodations for special needs students and English language learners. Plans are in place for disseminating information about HSEE to districts and schools immediately following a Board decision on adoption. There is not, however, a public document that lists all of the remaining steps with a timeline

for their completion. The development contractor's proposal covers many, but not all of these steps and there have been numerous updates to the timelines for the activities that are covered. We believe that wide review of plans and timeline for remaining development activities will help to identify potential issues and omissions in these plans while there is still time to address them.

Under current legislation, the Board must adopt an exit exam at its September 2000 meeting. The Board must subsequently review and approve implementing regulations and recommendations on detailed decisions such as how the minimum passing score is established and how test results are reported to students, parents, and schools. A more detailed timeline should inform the Board when required decisions are needed and when information for those decisions will be available. Others, including the HSEE Panel and districts and schools themselves, would similarly benefit from a clearer understanding of timelines and plans for remaining HSEE development and implementation activities.

We believe that a development timeline for HSEE would be most useful if it recognizes roles and tasks for as wide a range of stakeholder groups as possible, not just the test developer. Figure 5.1 lists a number of different groups with important roles in HSEE implementation and indicates some of the tasks each of these groups must complete. Figure 5.1 illustrates the range of activities that would be useful to include in a comprehensive timeline. In most cases, the indicated activities are well underway. This list is intended as an example only, not as a comprehensive list of timeline tasks.

In implementing the HSEE, California is setting ambitious goals for student achievement. These goals must be backed up with significant funding to help students achieve them. As suggested by the above list, the range of activities needed to prepare for HSEE implementation is quite broad. The Legislature has already approved funding for HSEE development and for programs to help districts prepare students for the HSEE, including funding for textbook adoption and teacher preparation. Additional resources will surely be needed for remedial programs and for other state and district efforts to ensure a fair opportunity for all students to pass the exam.

*Recommendation 3. CDE and the Board should work with districts to identify resource requirements associated with HSEE implementation. The Legislature must be ready to continue to fund activities to support the preparation of students to meet the ambitious challenges embodied in the HSEE.*

## Examples of Tasks for Inclusion in the HSEE Implementation Timelines

### For Districts and Schools:

1. Implement or improve instruction that covers the state content standards
2. Ensure that teachers are well prepared to deliver this instruction
3. Demonstrate that all students have access to this instruction
4. Provide notice to parents and students about HSEE requirement

### For the HSEE Panel:

1. Complete recommendations on accommodations and administration issues
2. Review essay questions with complete scoring guides and benchmark papers
3. Consider recommendations for scoring and reporting
4. Consider recommendations for setting minimum passing scores
5. Recommend materials describing test content and procedures
6. Recommend appropriate district/school preparation activities

### For CDE:

1. Work with the other stakeholders to maintain implementation timelines
2. Disseminate information to districts and work with them to prepare for HSEE
3. Monitor test development activities and facilitate communication on issues requiring decisions
4. Obtain legal advice on issues with HSEE implementation

### For the State Board:

1. Adopt blueprint specifying test content standards
2. Review and approve implementing regulations including recommendations for test accommodations, scoring, and reporting
3. Approve minimum passing scores

### For the Test Development Contractor:

1. Complete development and field testing of HSEE questions
2. Assemble test forms
3. Create sample test form and other test information/preparation materials
4. Develop detailed timelines for pre- and post-administration activities
5. Recommend/implement inclusion and accommodation policies
6. Recommend/implement procedures for recommending minimum scores
7. Recommend/implement procedures for scoring and reporting

### For the Independent Evaluators:

1. Review and comment on results from the fall field test
2. Review and comment on plans for testing, setting minimum passing scores, scoring, and score reporting
3. Collect data on preparations for HSEE and on the possible consequences of HSEE implementation

**Figure 5.1.** Examples of Items for Inclusion in the HSEE Development Timelines.

## Setting Achievable Standards

At its September meeting, the Board will be asked to approve the content to be assessed by the HSEE (content standards). At a subsequent meeting, the Board will be asked to approve recommendations on the minimum passing score (performance standard) needed to demonstrate adequate achievement relative to these content standards. One way to lessen risks associated with the current implementation schedule would be to ensure that the content and performance standards are not set so high that it will be difficult for most students to meet them.

There are two possible views of the general aims of the HSEE. One view is that California wants the HSEE to identify a moderately small number of students who are not reaching very minimum levels of achievement in mathematics and English/language arts and to help these students reach at least these minimum levels by the end of their high school years. An alternative view is that California wants the HSEE to motivate all students to achieve a broader range and higher levels of mathematics and language arts skills. These two views have very different implications for setting content and performance standards for the HSEE. This leads us to our next recommendation:

*Recommendation 4. The Board should adopt a clear statement of its intentions in setting HSEE content and performance standards. This statement should describe the extent to which these standards are targeted to ensure minimum achievement relative to current levels or to significantly advance overall expectations for student achievement.*

The content standards being recommended by the HSEE Panel appear to reflect the second view of HSEE goals. For example, a significant number of Algebra 1 standards are recommended while state law does not currently require algebra for graduation. It would appear that the proposed graduation standards are being used to drive changes toward a new curriculum rather than to identify minimal performers under the current curriculum.

Data reported in Chapter 3 above and in our June 30 report suggest that the average 10<sup>th</sup> grader from the Class of 2002 was likely to answer only about half of the HSEE questions correctly. While this information is tentative, and passing standards have not yet been established, these results suggest that it is likely that, even if the passing standard is as low as 50 % correct, roughly half of the students who take the test will fail on their first try<sup>6</sup>. There are several ways to reduce the probable failure rate:

- Develop easier questions. This is problematic because the HSEE Panel and others have consistently judged the current questions to be good measures of the standards they were designed to assess. Further, there is not time to develop and field-test new questions prior to the March 2001 operational administration.
- Set low standards for passing the exams. It is difficult to imagine setting a performance standard below the 50-percent-correct level. This would imply that

---

<sup>6</sup> The estimated passing rate is based on all 10<sup>th</sup> grade students. In 2001, 9<sup>th</sup> grade students will take the HSEE on a voluntary basis. Passing rates for this group may be significantly higher or lower depending on who volunteers.

students could pass the exam, even though they had not satisfactorily achieved more than half of the content standards. Note also that a frequent difficulty in setting minimum passing scores is that content experts expect the students to demonstrate at least a minimal level of achievement on every one of the content standards.

- Exclude content standards for which students are not yet well prepared. These standards could be introduced for later classes, after instruction related to them has been introduced or improved.

Unless one of the above approaches is taken, California appears to be trying to legislate the Lake Wobegon effect by requiring all students to be above average. Given better instruction and higher motivation, we would expect 10<sup>th</sup> graders in the Class of 2004 to perform somewhat better than 10<sup>th</sup> graders in the Class of 2002 did during the Spring 2000 field test. Requiring all students to score above a level that is likely to be higher than the Class of 2002 average would appear to be an unreasonable expectation relative to typical year-to-year gains in student performance. Yet a very plausible interpretation of current data and recommendations is that this is exactly where we are headed.

In 1980, President Bush and the nation's Governors established a set of National Educational Goals. Goal 5 was "By the year 2000, United States students will be first in the world in mathematics and science achievement." Meeting this goal would have required unprecedented gains in student achievement, and available evidence suggests this did not happen. Would a more plausible goal have been taken more seriously and resulted in greater gains? Unfortunately, there have not been controlled studies that would allow us to answer this question. Policy-makers face a difficult dilemma in attempting to set challenging goals that will maximize improvements in student achievement while avoiding expectations that are too unrealistic to be taken seriously. Parallels for the HSEE standards are obvious.

There should be a clear difference between setting goals for school-accountability where the focus tends to be on average achievement and setting goals for student-accountability that focus on the minimum acceptable level of achievement. To illustrate this difference, consider student grades in an Algebra 1 course. On average, we expect students to exhibit a level of performance that will earn them a C+ or B-, but students can pass the course with a minimum grade of D. Will the HSEE be targeting passing standards at the B- or D level? We offer the following recommendation consistent with the use of HSEE as a measure of minimum levels of student achievement:

*Recommendation 5. The Board should exhibit moderation in selecting content standards and setting performance standards for the initial implementation of HSEE. Standards should be subsequently expanded or increased based on evidence of improved instruction.*

We recognize that this recommendation may greatly increase the difficulty of meeting the current HSEE implementation schedule. If the Board makes significant changes to the proposed content standards at its September meeting, there are likely to be serious consequences for spring 2001 testing. There would not be time to develop and try out questions for any new content areas that are added. Fortunately, our recommendation for

moderation is aimed at limiting rather than expanding standards for minimum performance. If coverage is eliminated or reduced for some of the currently proposed standards, it would seem desirable to increase coverage of the remaining standards to maintain test score accuracy. Such an increase is likely to put a strain on the supply of available questions, but may still be feasible.

After the Board approves HSEE content standards, the contractor can assemble test forms for the spring 2001 administration that conform to these standards. Currently this is scheduled to happen about December of this year. At that point it will be possible to set performance standards, defined as minimum passing scores for each form. AIR is recommending a standard setting process for identifying a minimum score on a base form. This score level would then be equated to scores at the same level of difficulty on succeeding test forms. The “Bookmark” process being proposed is widely used in state assessment programs for setting achievement levels. We generally endorse the proposed approach, although we would like to see more detail on how and when consequence data (expected passing rates) will be introduced.

We believe that the panels charged with recommending performance standards should have access to the important considerations discussed in the setting of the content standards. For that reason, we offer the following recommendation for the process of setting HSEE performance standards.

*Recommendation 6. Members of the HSEE Panel and its Technical Advisory Committee should participate in developing recommendations for minimum performance standards.*

The Panel could also play a useful role in developing descriptions of the resulting content and performance standards for distribution to students, parents, and teachers. Many of the recommended standards are stated at a general level. It would be helpful to both students and the teachers charged with preparing the students to have a more extended description of each of these standards with examples and sample questions. Given limited time availability, the Panel’s role would be at most advisory, with detailed development work carried out by CDE and the development contractor.

### **Technical Suggestions for Improving the HSEE**

We have reviewed general summaries of AIR’s plans for conducting the fall field test, developing operational forms, and for spring 2001 operational administration activities. Selected development issues have been discussed with CDE’s Technical Studies Group. We are frankly concerned, however, that neither we, nor the HSEE Technical Committee, nor anyone else beyond the contractor’s staff, has had a chance to comment on technical issues in the contractor’s development plans. We think it would be highly advisable for these plans to be reviewed before they are implemented. In addition, given the real possibility of legal challenges to the HSEE, efforts to establish the defensibility of technical decisions would seem prudent. We offer a specific recommendation for obtaining additional technical review.

*Recommendation 7. CDE should move swiftly to establish an independent Technical Issues Committee (TIC) to recommend approval or changes to the HSEE development*



*contractor's plans for item screening, form assembly, form equating, and scoring and reporting.*

CDE has already initiated efforts to obtain independent experts to review technical issues in implementing the HSEE. This would be a significant improvement over the current situation, and could fully address our concern with the need for independent review of technical issues. There is a need for timely advice and a large committee might make take too long to reach consensus. The group that we envision would differ from current CDE efforts by including one or two additional technical reviewers with complementary expertise.

An example may illustrate one need for a broad consensus on technical issues in HSEE development. Field test plans were developed without a clear review of item analysis and screening procedures. Sample sizes were adequate for the overall screening and calibration of the test questions. However, no effort was made to emphasize minority students in sampling and, as a consequence, sample sizes for detecting differential item functioning for African American students were marginal for multiple choice questions and completely inadequate for the essay questions. The state could be legally vulnerable if scores are used operationally to fail students without an adequate check for differential functioning across key demographic groups. It should be feasible to use data from the operational administration of HSEE forms to check for inappropriate group differences. So long as no problems are found, this might not impact the reporting schedule significantly. An independent review of technical aspects of the field test plans might have improved the design considerably and eliminated the need for important analyses during the initial operational score reporting cycle. More importantly, such review can still improve the process as it continues toward full implementation.

A related concern is that very limited information will be available on possible passing rates as the minimum passing score is being established. Approximate information can be pieced together from the field test results on an question-by-question basis, but there will not be much basis for estimating, in advance, the degree of adverse impact for some minority groups (e.g., African Americans) resulting from alternative passing scores.

One of the requirements for the independent evaluation is to provide recommendations for improving the HSEE. Given our reporting schedule, we cannot provide timely advice on all of the technical issues that must be addressed. We are recommending one or more independent reviewers who can be available to address issues quickly as they arise. In the remainder of this section, we do offer advice on two technical issues of immediate concern. The issues are constructing test forms so that comparable scores can be obtained (test equating) and providing feedback to schools that participate in the field tests.

Equating scores from different test forms. Plans for equating different test forms need to be reviewed and finalized before the initial forms are assembled. The contractor's current plan appears to rely on a "pre-equating" approach, where item statistics from the two field tests provide the only basis for creating comparable (equated) scores. We do not believe that total reliance on this form of equating will be sufficient. There are plenty of examples of item position and context effects that would not be accounted for under this plan. It is also likely that differences in student motivation between the field test and operational administrations will affect different questions differently.

In many programs, a common set of “anchor” questions is included in two or more different forms to provide a basis for estimating differences in performance on these forms. For example, achievement differences between the March and May 2001 examinees could be assessed on the basis of a common set of questions. These estimated differences would then be compared to the differences estimated from all of the test questions, based on the “pre-equated” scores and adjustments introduced if needed.

It may not be feasible to use some of the same scored questions in both the March and May forms because of test security concerns or because of a desire to release all of the test questions from the initial administration. AIR’s current plan calls for development of a number of different versions of each operational test form with a different set of about 20 non-operational (unscored) questions included in each version. These questions are being administered as a field test and would not count in determining each student’s operational score. We recommend that one or two additional versions be developed with a common set of previously screened (all good) questions in the unscored positions. The same set(s) of questions could also be included as unscored questions in the May 2001 test form, providing a basis for linking (or checking the linkage of) scores across these administrations.

Feedback to schools participating in the Field Test. Given the amount of operational testing that exists in schools today, it is understandably difficult to recruit schools to participate in non-operational testing, such as the HSEE field tests. Currently there are no plans to provide any information to participating schools on the performance of their students in the field test. The field test booklets include questions that will be screened out and they are not operational forms. In addition, passing standards have not yet been set. While it is not possible to provide reliable information on individual students, it may still be possible to provide some information to schools. A large number of questions are administered in the field test, roughly 400 for each subject rather than only 100 in an operational administration. It should be possible to provide some information on relative passing rates for questions assessing each specific content standard, pooling data across both students and the questions that assess these standards. The tables in Chapter 3 of this report show passing rates for the questions under each content standard from the spring 2000 field test. It should be possible to construct similar tables showing passing rates by content standard for a particular school in comparison to passing rates for the state as a whole. We believe that such information would be highly valuable to schools in identifying areas where additional preparation is needed. Offering to provide such information could be a strong incentive for participation in the field test.

### **Other Recommendations in the June 30 Evaluation Report**

Clarifying Relationships Among State High School Testing Programs. Now that some scores for the STAR standards-based exam have been reported, it seems more important than ever to find a basis for comparing them to HSEE scores. Schools will want to know how performance for individual students on one exam predicts performance on the other. Indeed, proposals have been discussed for exempting students from the HSEE on the basis of scores on the standards-based STAR assessment or scores from the Golden State exams. We had recommended linking STAR and HSEE scores for specific students, but there are some issues with the feasibility of this approach due to difficulties in identifying particular

students. There are many ways the scores from the two assessments might be linked. One alternative would be to administer some of the standards-based STAR questions to HSEE examinees, perhaps as non-scored questions. The information obtained would help in ensuring consistency in standards and expectations for the two assessments.

Dissemination and Assistance to Local Districts. Our June 30 report contained a recommendation on the importance of dissemination and assistance for local districts. CDE has worked to develop dissemination plans. Full implementation of these plans awaits the Board's approval of the examination in September. We have nothing further to add at this time, except to say that funding issues with respect to dissemination and assistance activities should definitely be addressed under Recommendation 3 above.

Appropriate Accommodations. The HSEE Panel has collected a great deal of information on options for accommodating special needs students and English language learners. Specific recommendations were discussed at their July 2000 meeting and will be presented to the Board in September. More work is clearly needed to prepare for the possibility that the state or individual districts may be sued by students who feel that they were inappropriately included or excluded or did not receive a comparable score due to some issue with accommodations.

